AD-A262 789

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

2. REPORT DATE
February 1993

3. REPORT TYPE AND DATES COVERED
Professional Paper

4. TITLE AND SUBTITLE
ON MODELING OF IF-THEN RULES FOR PROBABILISTIC INFERENCE

5. FUNDING NUMBERS
PR: ZW40
PE: 0601152N
WU: DN300173

6. AUTHOR(S)
I. R. Goodman and H. T. Nguyen

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Command, Control and Ocean Surveillance Center (NCCOSC)
RDT&E Division
San Diego, CA 92152–5001

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Chief of Naval Research
Independent Research Programs (IR)
OCNR-10P
Arlington, VA 22217–5000

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

We identify various situations in probabilistic intelligent systems in which conditionals (rules) as mathematical entities as well as their conditional logic operations are needed. In discussing Bayesian updating procedure and belief function construction, we provide a new method for modeling if ... then rules as Boolean elements, and yet, compatible with conditional probability quantifications.

To be published in *Journal of Intelligent Systems*.

14. SUBJECT TERMS
data fusion
conditional events
syntactics
semantics
linguistic information

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAME AS REPORT

21a. NAME OF RESPONSIBLE INDIVIDUAL	21b. TELEPHONE (Include Area Code)	21c. OFFICE SYMBOL
I. R. Goodman	(619) 553-4014	Code 421

ON MODELING OF IF-THEN RULES FOR PROBABILISTIC INFERENCE

HUNG T. NGUYEN
DEPARTMENT OF MATHEMATICAL SCIENCES
NEW MEXICO STATE UNIVERSITY
Las Cruces, NM 88003

AND

I.R. GOODMAN
ASHORE COMMAND & INTELLIGENCE
CENTERS DIVISION
CODE 421, NRaD
San Diego, CA 92152-5000

ACCEPTED FOR PUBLICATION IN THE
JOURNAL OF INTELLIGENT SYSTEMS

ON MODELING OF IF-THEN RULES FOR PROBABILISTIC INFERENCE

HUNG T. NGUYEN
DEPARTMENT OF MATHEMATICAL SCIENCES
NEW MEXICO STATE UNIVERSITY
Las Cruces, NM 88003

AND

I.R. GOODMAN
ASHORE COMMAND & INTELLIGENCE
CENTERS DIVISION
CODE 421, NRaD
San Diego, CA 92152-5000

Abstract. We identify various situations in probabilistic intelligent systems in which conditionals (rules) as mathematical entities as well as their conditional logic operations are needed. In discussing Bayesian updating procedure and belief function construction, we provide a new method for modeling if ... then rules as Boolean elements, and yet, compatible with conditional probability quantifications.

Key words. Bayesian updating, belief functions, conditionals, probabilistic inference.

1 Introduction

In probabilistic systems, the production rules (if ... then ... rules) connect events and are quantified by conditional probabilities. With additional structures, such as conditional independence, the problem seems feasible and

computations are based entirely on the standard calculus of probabilities (see e.g. Pearl, 1988).

The situation is far from clear when events of interest are conditional events. In this paper, we will point out situations in which these problems occur. When we try to extend probabilistic techniques to these situations, we realize that new objects and tools are needed. It all boils down to modeling if ... then ... rules in some appropriate fashion, and yet compatible with conditional probability evaluations.

2 Why do we need a mathematical concept of conditional events?

To be specific, propositions or events are viewed as elements of a σ-algebra $\mathcal{A}$ of subsets of a universe of discourse Ω. The pair $(\Omega, \mathcal{A})$ thus denotes a measurable space. We use letters a, b, c ... to denote elements of $\mathcal{A}$. Set operations are: $\wedge$ (or simply $\cdot$, for intersection), $\vee$ (union), $(\cdot)'$ (complement), $\leq$ (set-inclusion), $\emptyset$ (empty set).

There is more than one way to quantify a rule of the form "if b then a" by probabilities. In the context of two-valued logic, this rule, symbolized as $b \rightarrow a$, is interpreted as, material implication, that is $b \rightarrow a = b' \vee a$, which is an element of $\mathcal{A}$. If P is a probability measure on $\mathcal{A}$, then the strength of the rule $b \rightarrow a$ can be taken as $P(b' \vee a)$. See e.g. Nilsson, 1986.

However, due to the meaning, as well as to the uncertainty involved, the quantification of $b \rightarrow a$ is via conditional probability, that is $P(b \rightarrow a) = P(a \mid b)$, provided $P(b) > 0$. If we take this viewpoint, then $b \rightarrow a$ cannot be modeled by material implication, since $P(a \mid b) \neq P(b' \vee a)$, in general. More importantly, if $b \rightarrow a$ is quantified by $P(a \mid b)$, then $b \rightarrow a$ cannot be an element of $\mathcal{A}$. This is known as Lewis' triviality result (Lewis, 1976). In probabilistic systems (see e.g. Pearl, 1988), the modeling of causal relationships among variables of interest (in some knowledge domain) seems unnecessary. That is, one does not need to define $b \rightarrow a$ as some mathematical entity. In contrast, relations among variables, such as conditional independence, and the assignment of conditional probabilities to rules (expressed in a natural language), as well as prior probabilities, suffice to specify a joint probability distribution on all variables involved, so that probabilistic inference can

be carried out. This is somewhat similar to situations in probability and statistics: the quantity $P(a \mid b)$ stands for $P_b(a) = P(ab)/P(b)$ where P_b is a probability measure on $\mathcal{A}$, defined as $P_b(a) = P(ab)/P(b)$. Although, DeFinetti (1974) did consider $(a \mid b)$ as a mathematical entity, namely an object with three "truth"-values: true (when both a and b are true), false (when a is false and b is true) and undetermined (when b is false), this observation does not contribute anything new to probability and statistics. It is interesting to point out that, in the same vein, as far as we know, the concept of "conditional random variables" was mentioned only in Wilks (1963), in an intuitive setting.

The common point is this. While one is free to ask questions and pursue mathematical investigations, the results obtained will be marginal and hence ignored if they do not lead to advances in applications. See Goodman, Nguyen and Walker (1992) for a history of the mathematical investigations of conditional events.

As we will see, it turns out that the need to model conditional events or production rules as mathematical entities (as opposed to primitives in natural languages, as in Adams, 1975, or in the general discussions in the philosophical community) is apparent in the field of expert systems where, adopting Bayesian methodology, one insists on using probabilistic techniques for the management of uncertainty. This is essentially due to the fact that, intelligent systems are concerned with reasoning with knowledge. Now, not only knowledge can be represented in different forms, but it is, in general, expressed in some conditional form.

In the following, we will illustrate the above need. Recall that we write $b \rightarrow a$ for "if b then a", and use $P(a \mid b)$ to specify the strength of this rule.

(i) This example is inspired from Adams (1992). Consider a box containing red, blue and white balls with unknown proportions. We are interested in the probability of getting a blue ball on the first drawn of a ball from this box. Suppose that we learn the information "there are many more blue balls than white balls" (or even with more precise numerical information, such as $P(\text{blue}|\text{not red}) \approx .99$). Let us examine the heuristic expression $P(\text{blue}|(\text{blue}|\text{not red}))$.

As emphasized in Adams (1992), the above expression cannot be written in standard probability theory, since the antecedent (blue|not red) (or not red$\rightarrow$blue) is not yet defined mathematically, and more over, as mentioned earlier, even if it can be defined, it does not belong to the domain of P. Thus,

first we need to model (not red→blue), and then (not red→blue)→blue as an iterated conditional. Once this task is completed, we still have to specify an associated probability measure on the new space of conditionals to give a rigorous formulation of the above heuristic expression.

Let us pursue this example a little further. In expert systems, we usually have several rules, say "$b_i \to a_i$", $i = 1, 2, \ldots, n$. To evaluate the probability of some event of interest c from this rule-base, we formally write

$$P(a \mid (b_1 \to a_1) \text{ and } (b_2 \to a_2) \text{ and } \ldots (b_n \to a_n)).$$

The combination of rules, say, via the logical connective "and" can be carried out if "and" is specified. This is basically the problem of "reasoning with conditional knowledge", in which we need to specify a logic, that is, an algebraic structure of conditionals.

(ii) A basic inference principle in rule-based systems is Modus ponens.

In two-valued logic framework, where $b \to a = b' \vee a$, we deduce a if the evidence b holds. This is because here the partial order $\leq$ (set inclusion) is precisely the entailment relation. Specifically:

$$(b \to a)b = (b' \vee a)b = ab \leq a.$$

When the evidence $c \neq b$, one can obtain a degree of uncertainty on a by computing $P[(b' \vee a)c]$.

Similarly, in fuzzy logic (see e.g. Yager et al, 1987), extending classical logic, and where a, b, c become fuzzy sets, the conclusion of modus ponens takes the form $(b \to a)c$, describing a new fuzzy set in which "conjunction" is chosen as some t-norm (e.g. minimum operator), and the fuzzy implication $b \to a$ is interpreted using some truth table for the fuzzy implication (binary) operator $\to$.

However, if we insist on the quantification $P(b \to a) = P(a \mid b)$, we have to proceed differently, again, since $b \to a$ will be no longer $b' \vee a$.

Writing $b \to a$ as a conditional object $(a \mid b)$, and identifying c with $(c \mid \Omega)$, we can form $(a \mid b)(c \mid \Omega)$ where conjunction of conditionals need to be specified. From that, a computation of probability is possible.

(iii) As mentioned in Goldszmidt and Pearl (1992), the ruled-base of an expert system might contain a rule of the form "If $(b \to a)$ then $(d \to c)$", symbolized as $(a \mid b) \Rightarrow (c \mid d)$.

It is obvious that to quantify this rule by conditional probability, i.e. computing $P[(a \mid b) \Rightarrow (c \mid d)]$, we first have to define the objects like $(a \mid b)$. Next, $\Rightarrow$ is some implication operator among these conditional objects, which can be derived from logical operations among these objects. Finally, what probability measure on the space of conditionals to use in order to quantify the rule?

3 Bayesian updating and belief construction

Before going into probabilistic inferences such as Bayesian updating procedures and combination of evidence using belief functions (Shafer, 1976), let us outline briefly previous efforts on formulating a mathematical theory of conditionals (see e.g. Goodman, Nguyen and Walker, 1991, for details).

Consider again a measurable space $(\Omega, \mathcal{A})$. In view of Lewis' triviality result, there is no binary operation $\rightarrow$ from $\mathcal{A} \times \mathcal{A}$ to $\mathcal{A}$ (where $\times$ denotes cartesian product) such that for any a, $b \in \mathcal{A}$, and any probability measure P on $\mathcal{A}$ with $P(b) > 0$, one has

$$P(b \rightarrow a) = P(a \mid b).$$

Thus, in modeling the rule $b \rightarrow a$, whose quantification is $P(a \mid b)$, one has to go "outside" of $\mathcal{A}$.

One axiomatic derivation leads to a representation of $b \rightarrow a$ as an "interval" in $\mathcal{A}$ (see also, Nguyen, 1992), namely

$$\begin{aligned} b \rightarrow a &= \{x \in \mathcal{A} : ab \leq x \leq b' \vee a\} \\ &= [ab, b' \vee a] \text{ for short.} \end{aligned}$$

When $a = b$, by identifying $[a, a]$ with a, we see that in general, $b \rightarrow a$ lies outside of $\mathcal{A}$.

It is easy to check that $[a, b] = b' \vee a \rightarrow a$ (since $a \leq b$), hence the space of all closed intervals, denoted as $\mathcal{A} \mid \mathcal{A}$, is precisely that of all conditionals $b \rightarrow a$.

This space contains $\mathcal{A}$ strictly. Contrary to a statement in Gilio and Spezzaferri (1992), these conditionals are equivalent to DeFinetti's conditional events. To see this, viewing Ω as the set of all models in a logical

setting, DeFinetti's conditional event $(a \mid b)$ is identified with the generalized indicator function (see Schay, 1968)

$$(a \mid b)(\omega) \quad = \begin{cases} 1 \text{ for } \omega \in ab \\ 0 \text{ for } \omega \in a'b \\ u \text{ for } \omega \in b'. \end{cases}$$

It is obvious that such functions are in one-to-one correspondence with elements of $\mathcal{A} \mid \mathcal{A}$, since they specify b' and ab, and conversely.

In fact, it is precisely this three-valued logic connection that one can discover all possible algebraic structures of $\mathcal{A} \mid \mathcal{A}$. For example, Lukasiewicz' three-valued logic (see e.g. Rescher, 1969) will equip $\mathcal{A} \mid \mathcal{A}$ with interval operations. That is,

$$[a,b] \wedge [c,d] = [ac, bd]$$

$$[a,b] \vee [c,d] = [a \vee c, b \vee d].$$

Note that $\mathcal{A} \mid \mathcal{A}$ is not a Boolean algebra since it is not complemented. Indeed, if $a \leq b$ then $b' \leq a'$.

However, this bounded, distributive lattice has a pseudo-complementation:

$$[a,b]^* = [b', b'],$$

satisfying Stone's identity

$$[a,b]^* \vee [a,b]^{**} = [1,1]$$

so that $\mathcal{A} \mid \mathcal{A}$ is a Stone algebra (see e.g. Gratzer, 1978).

The above investigations provide a new mathematical framework for manipulating conditional information.

While the mathematical concept of a conditional event, or of a production rule, is well understood, one would like also to consider some other equivalent representation of $b \to a$ which possesses some "boolean" flavor. This would be useful as in the following situations.

(i) Suppose that P is a prior probability measure on $(\Omega, \mathcal{A})$. When we learn that some event $a \in \mathcal{A}$ has occurred, we update our knowledge P by conditioning on a, that is, change P to P_a. How can we continue to do so if, instead of learning a, we learn a rule $b \to a$? Viewing $b \to a$ as a conditional event $[ab, b' \vee a]$, how do we make sense of $P_{[ab,b' \vee a]}$ as a new probability

measure? The difficulty seems to lie in the fact that $[ab, b' \vee a] \in \mathcal{A} \mid \mathcal{A}$ which is not a Boolean σ-algebra.

(ii) In the simplest situation of using belief functions to quantify our degrees of belief (see Shafer, 1976), one can construct a belief function F on Ω (finite) from the knowledge of $P(a)$, for some subset a of Ω as follows. Define the assignment mass function $m : \mathcal{P}(\Omega) \to [0, 1]$, where $\mathcal{P}(\Omega)$ denotes the power set of Ω, by $x \in \mathcal{P}(\Omega)$,

$$m(x) = \begin{cases} P(a) & \text{if } x = a \\ 1 - P(a) & \text{if } x = \Omega \\ 0 & \text{otherwise.} \end{cases}$$

And then, as usual, for $y \in \mathcal{P}(\Omega)$,

$$F(y) = \sum_{x \leq y} m(x).$$

As in the Bayesian updating case, suppose we know a, b and $P(a \mid b)$, how should we proceed to construct an associated assignment mass function?

The difficulty is similar to that in the Bayesian updating case.

In view of situations as above, we are going to investigate, in the next section, a "booleanization" of conditionals which should provide a new tool for probabilistic inference with conditional information.

4 A booleanization of conditionals

Recall that a rule of the form $b \to a$ cannot be modeled as an element of the σ-algebra $\mathcal{A}$, as long as we want to quantify it by $P(b \to a) = P(a \mid b)$, for any probability measure P on the measurable space $(\Omega, \mathcal{A})$.

In Section 3, we mentioned the conditional space $\mathcal{A} \mid \mathcal{A}$, strictly larger than $\mathcal{A}$, which admits $b \to a$ in its elements. However, $\mathcal{A} \mid \mathcal{A}$ is not a σ-algebra. We are going to search for a σ-algebra larger than $\mathcal{A} \mid \mathcal{A}$ for which rules $b \to a$ are its elements.

We start from the following remark of D. Bamber, NRaD (personal communication),

$$P(a \mid b) = P(ab)/P(b) = \frac{P(ab)}{1 - P(b')} = \sum_{n=0}^{+\infty} P(ab)[P(b')]^n$$

(using $\frac{1}{1-r} = \sum_{n=0}^{+\infty} r^n$, for $0 \leq r < 1$). The term $[P(b')]^n$ suggests a product measure of the set $b' \times b' \times \cdots \times b'$ (n times), where $\times$ denotes cartesian product.

Since n runs over the set of non-negative integers, an infinite (countable) product space is required.

Thus, let $\widehat{\Omega}$ be the infinite cartesian product of Ω, i.e.

$$\widehat{\omega} \in \widehat{\Omega},\ \widehat{\omega} = (\omega_1, \omega_2, \omega_3, \ldots),\ \omega_n \in \Omega,\ n \geq 1.$$

A cylinder in $\widehat{\Omega}$ is a subset of $\widehat{\Omega}$ of the form $a_1 \times a_2 \times \cdots \times a_n \times \Omega \times \Omega \times \ldots$, for $n \geq 1$, and $a_i \in \mathcal{A}$, $i = 1, \ldots, n$.

To simplify notation, we write $a_1 \times a_2 \times \cdots \times a_n$ to mean the cylinder with this base. Thus, for example, ab is viewed as the cylinder $ab \times \Omega \times \Omega \cdots$, and $b' \times b'$ is the cylinder $b' \times b' \times \Omega \times \Omega \times \cdots$, and so on.

Let $\widehat{\mathcal{A}}$ be the infinite product σ-algebra on $\widehat{\Omega}$, that is, the smallest σ-algebra containing all cylinders of $\widehat{\Omega}$.

Let $\widehat{P}$ denote the product measure on $(\widehat{\Omega}, \widehat{\mathcal{A}})$ with identical one-dimensional marginals P, that is

$$\widehat{P}(a_1 \times a_2 \times \cdots \times a_n \times \Omega \times \Omega \times \cdots) = P(a_1)P(a_2)\ldots P(a_n),\ \forall n \geq 1.$$

Now, observe that the cylinders ab, $b' \times ab$, $b' \times b' \times ab, \ldots$ are pairwise disjoint in $\widehat{\Omega}$. Indeed

$$ab = ab \times \Omega \times \Omega \times \cdots = \{\widehat{\omega} = (\omega_1, \omega_2, \omega_3, \ldots) : \omega_1 \in ab, \omega_n \in \Omega, n \geq 2\}$$

$$b' \times ab = \{\widehat{\omega} = (\omega_1, \omega_2, \omega_3, \ldots) : \omega_1 \in b', \omega_2 \in ab, \omega_n \in \Omega, n \geq 3\},$$

(note that ab and $ab \times b'$ are not disjoint).

Consider the map

$$f : \mathcal{A} \times \mathcal{A} \rightarrow \widehat{\mathcal{A}}$$

defined by

$$f(a, b) = ab \vee (b' \times ab) \vee (b' \times b' \times ab) \vee \ldots$$

where, by abuse of notation, $\vee$ stands for set union in $\widehat{\Omega}$. Note that $f(a, b)$ is a countable union of cylinders, and hence is an element of $\widehat{\mathcal{A}}$.

Since $\widehat{P}$ is a probability measure, we have

$$\begin{aligned}\widehat{P}(f(a,b)) &= \widehat{P}[ab \vee (b' \times ab) \vee (b' \times b' \times ab) \vee \ldots] \\ &= \widehat{P}(ab) + \widehat{P}(b' \times ab) + \widehat{P}(b' \times b' \times ab) + \ldots \\ &= P(ab) + P(b')P(ab) + P(b')P(b')P(ab) + \ldots \\ &= \sum_{n=0}^{+\infty} P(ab)[P(b')]^n = P(a \mid b).\end{aligned}$$

Thus, the probability space $(\widehat{\Omega}, \widehat{\mathcal{A}}, \widehat{P})$ extends $(\Omega, \mathcal{A}, P)$ in the sense that, for $a \in \mathcal{A}$, $\widehat{P}(f(a,\Omega)) = P(a)$, and for a, $b \in \mathcal{A}$ with $P(b) > 0$, $\widehat{P}(f(a,b)) = P(a \mid b)$.

In view of this matching, the rule $b \rightarrow a$ can be modeled as $f(a,b)$ which is an event, but in another measurable space.

Now, given $b \rightarrow a$, we can update P rigorously by $P_{b \rightarrow a}$, as in the unconditional information case. Indeed, we take $P_{b \rightarrow a}$ to be $\widehat{P}_{f(a,b)} : \widehat{\mathcal{A}} \rightarrow [0,1]$, which is a usual conditional probability measure: For

$$A \in \widehat{\mathcal{A}},\ \widehat{P}_{f(a,b)}(A) = \widehat{P}[A \wedge f(a,b)]/\widehat{P}(f(a,b))$$

where $\wedge$ stands for set intersection in $\widehat{\Omega}$. For $c \in \mathcal{A}$, we take $P_{b \rightarrow a}(c) = \widehat{P}_{f(a,b)}(c \times \Omega \times \Omega \times \cdots)$.

As a final remark, while the booleanization of conditionals provide a rigorous framework for probabilistic inference when dealing with conditional information, the computations might be complicated. It is anticipated that logical operations among conditionals (viewed as intervals in Boolean σ-algebras) can be used as approximations to Boolean operations on $\widehat{\Omega}$, and thus reduce the complexity of computational problems.

Acknowledgements. We express out thanks to D. Bamber, P. Calabrese and E. Walker for their constructive discussions concerning this work.

References

[1] Adams, E. (1975). *The Logic of Conditionals.* D. Reidel, Dordrecht, The Netherlands.

[2] Adams, E. (1992). Conditional information from the point of view of the logic of high probability. Preprint.

[3] DeFinetti, B. (1974). *Theory of Probability.* J. Wiley, N.Y.

[4] Gilio, A. and Spezzaferri, F. (1992). Knowledge integration for conditional probability assessments. *Proceedings 8th Conf. Uncertainty in AI*, Stanford Univ., July 1992, Morgan Kaufmann, 98-103.

[5] Goldszmidt, M. and Pearl, J. (1992). Reasoning with qualitative probabilities can be tractable. *Proceedings 8th Conf. Uncertainty in AI*, Stanford Univ., July 1992, Morgan Kaufmann, 112-120.

[6] Goodman, I.R., Nguyen, H.T. and Walker, E.A. (1991). *Conditional Inference and Logic for Intelligent Systems.* North Holland, Amsterdam.

[7] Gratzer, G. (1968). *General Lattice Theory.* Birkhauser, Basel.

[8] Lewis, D. (1976). Probabilities of conditionals and conditional probabilities. *Phil. Rev.* (85), 297-315.

[9] Nguyen, H.T. (1992). Intervals in boolean rings: Approximation and logic. To appear in *J. Foundations of Computing and Decision Sciences.*

[10] Nilsson, N. (1986). Probability logic. *Art. Intell.* (28), 71-87.

[11] Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems.* Morgan Kaufmann.

[12] Schay, G. (1968). An algebra of conditional events. *J. Math. Anal. and Appl.* (24), 334-344.

[13] Shafer, G. (1976). *A Mathematical Theory of Evidence.* Princeton Univ. Press.

[14] Wilks, S. (1963). *Mathematical Statistics.* J. Wiley.

[15] Yager, R.R., Ovchinnikov, S., Tong, R.M., and Nguyen, H.T. (1987). *Fuzzy Sets and Applications: Selected Papers by L.A. Zadeh.* J. Wiley, N.Y.